

## INTEGRATING GEO-STATISTICS AND GEOGRAPHIC INFORMATION SYSTEMS IN GEOTECHNICAL SOIL PROFILING

ENG. MOHAMMED ABD EL FADIL ISMAIL, MSc. , REGISTERED TO PHD,  
Department of Structural Engineering, Ain Shams University.

PROF. DR. ALI ABD EL FATAH, PROFESSOR OF GEOTECHNICAL ENGINEERING,  
DEPARTMENT OF STRUCTURAL ENGINEERING, AIN SHAMS UNIVERSITY

DR. ASHRAF M. HEFNY, ASSISTANT PROFESSOR OF GEOTECHNICAL ENGINEERING,  
DEPARTMENT OF STRUCTURAL ENGINEERING, AIN SHAMS UNIVERSITY

### Abstract

Since its evolution, Geotechnical engineering is considered one of the sciences that deeply based on uncertainties. These uncertainties are attributed to two facts. First fact, the formulas used in geotechnical engineering are empirical formulas and could be deeply affected by the non-homogeneity of soil formation. The second fact is that always the soil is being characterized through a group of boreholes that are considered representative for the whole site, which could result in the missing of some anomalies in soil composition. Hence, geo-technical engineers are always looking for a technique, that can be used in dealing with these uncertainties. Several techniques were adopted starting from the sixties of the previous century like knowledge base expert systems, finite element analysis and finally geo-statistics. But the most sidelined of these techniques was geo-statistics due to several reasons. Hence, this paper tries to introduce geo-statistics focusing on its applications in geotechnical engineering specially in soil profiling. It discusses the reason behind the fact that still the actual implementation of Geo-statistics in Geotechnical engineering is without real presence in the traditional work of any geotechnical engineer. Moreover, it introduces a look ahead towards actual implementation of Geo-statistics in soil profiling through the integration between one of the most mature geo-statistical software (Sgems) and the GIS software which is ARCGIS of ESRI

تعتبر الهندسة الجيوتقنية منذ تطورها بمجهودات العديد من المهندسين و علماء الرياضيات و الاحصاء أحد العلوم التي ترتبط بقوة بأوجه عدم اليقين، و يرجع هذا لسببين اساسيين، و هما أن كافة الصيغ و المعادلات المستخدمة في الهندسة الجيوتقنية هي معادلات تجريبية تعتمد بشكل كبير على اساليب الميكانيكا الكمية و تلك المعادلات يمكن أن تتأثر بشدة بحقيقة عدم تجانس التربة اثناء تكوينها في الحقب الجيولوجية المختلفة مما يؤدي الى وجوب اعتبار الاحتمالات المختلفة عند نمذجة التربة، و الحقيقة الثانية تعود للأسلوب المستخدم لنمذجة التربة من خلال مجموعة من الجسات بالحفر أو سجلات بالتجارب الحقلية مثل CPT مع اعتبار ان تلك الجسات أو السجلات هي ممثل لمجسم التربة بالكامل مما قد يؤدي الى فقد العديد من المعلومات الخاصة بالتربة في حالة التكوينات الغير نمطية، و من ثم كان أحد أهداف علماء الهندسة الجيوتقنية دائما هو ايجاد تقنيات للتعامل مع واقع عدم اليقين المتعلق بدراسات التربة، و قد ظهرت بداية من ستينيات القرن الماضي العديد من التقنيات التي استخدمت لهذا الغرض و منها تقنيات الانظمة الخبيرة و تقنيات الانظمة الغير محدودة و أخيرا تقنيات الاحصاء المكاني و التي تعتبر هي التقنية التي عانت من التهميش نسبيا في هذا الغرض و ذلك للعديد من الاسباب، و من ثم فإن تلك الورقة تحاول عرض تقنيات الاحصاء المكاني و تركز على تطبيقاتها في الهندسة الجيوتقنية و بالاصح عمل النموذج الثلاثي الابعاد لمجسم التربة، كما تناقش اسباب حقيقة ان الاحصاء المكاني مازال حبيس الورقات العلمية دون وجود واقع تطبيقي فعلي في الحياة اليومية لمهندسي الجيوتقنية التقليديين، علاوة على ذلك في تقدم نظرة مستقبلية نحو تفعيل تلك الانظمة في الهندسة الجيوتقنية من خلال تجربة عملية للتكامل بين أحد اشهر برامج نظم المعلومات الجغرافية ArcGIS و بين أحد برامج الاحصاء المكاني ثلاثي الابعاد الاكثر نضجا عالميا وفق تقييمات الخبراء بهذا المجال وهو Sgems و الخاص بجامعة Stanford

**Keywords:** Geo-statistics, GIS, Sgems, ArcGIS, variogram.

### 1. Introduction

Until the 18th century soil characterization and behavior prediction were considered arts more than a science. There was not any theoretical base for

behavior analysis and it was depending on previous experience. [3].The problem in this orientation was the lack of documented previous experiences and the unavailability to exchange experiences to match soil

patterns which is not the case after the evolution of electronic computing and huge data banks in the sixties of the 20<sup>th</sup> century. This problem drove engineers and scientists to standardize some formulas and equations. One of their trials to enhance the geotechnical modeling process was through the utilization of the techniques of computational mechanics like finite element, finite difference, finite volume and boundary element methods [4]. But these methods were not able to account for some geotechnical problems like soil characterization. Moreover all these techniques could not limit the dependence on assumptions to define the properties in the media of analysis and even the governing equation that relates the property of the medium of analysis and the applied actions with the expected behavior. [4] In case of geotechnical engineering the computational mechanics techniques treated soil as an elastic homogeneous isotropic medium which is an abstract assumption that increase the uncertainty in the results. With the continual development of computers and storage media, engineers started to bounce back to the concept of using past experience in geotechnical analysis and to utilize the concepts of experimental and applied soil mechanics instead of theoretical and computational ones. This orientation was supported by the appearance of Artificial intelligence (AI) techniques (eg. data mining, knowledgebase systems, expert systems) and also the appearance of geo-statistical estimations and simulations. In both techniques (AI and geo-statistics) uncertainty is considered. In knowledgebase systems for example, uncertainty is considered through the usage of Bayesian or probabilistic inference [5] while in geo-statistics it is considered through Variance surfaces which are one of the outputs of geo-statistical modeling [6]. Although AI and geo-statistics techniques were synchronous in their evolution, the opportunity of AI to participate in geotechnical modeling was greater. This fact is attributed to several reason, one of which is the early development of geotechnical software that utilizes AI expert system principals like LOGOS and SiteCHAR [7] which is not the case of geo-statistics. Moreover, the existing mature geo-statistical software lacks many spatial querying, attribute querying, presentation and reporting tools that are needed for any Geotechnical process modeling. While some other geo-statistical software modules are hidden within mining flavored programs . Beside that still the geo-statistical parameters and algorithms are not well known for geotechnical engineers. [12]

## 2. What is geo-statistics

In 1950's Georges Matheron recognized the approaches of the geologist D.G Krige in spatially correlating large amount of data in predicting a real

gold concentrations in South Africa. In 1965, in the Ecole des Mines, France, Matheron was able to formulate the equations and algorithms of what is currently known as Geo-statistics and he honors the role of D.G krige by using the term "Kriging" to describe the process of estimating the values at un-sampled locations. [8]

Geo-statistics could be defined as "Statistical technique offers a way of describing the spatial continuity of natural phenomena and provides adaptations of classical regression techniques to take advantage of this continuity." [9] It depends on Topley first law of geography "Things that are closer together tends to be more alike than things that are far apart". It assumes that all values in the study area are the result of a random process with dependence. In a spatial or temporal context, such dependence is called autocorrelation. This concept was abstracted through an equation:

$$Z(s) = \mu(s) + \varepsilon(s) + m \quad [6] \quad .[1]$$

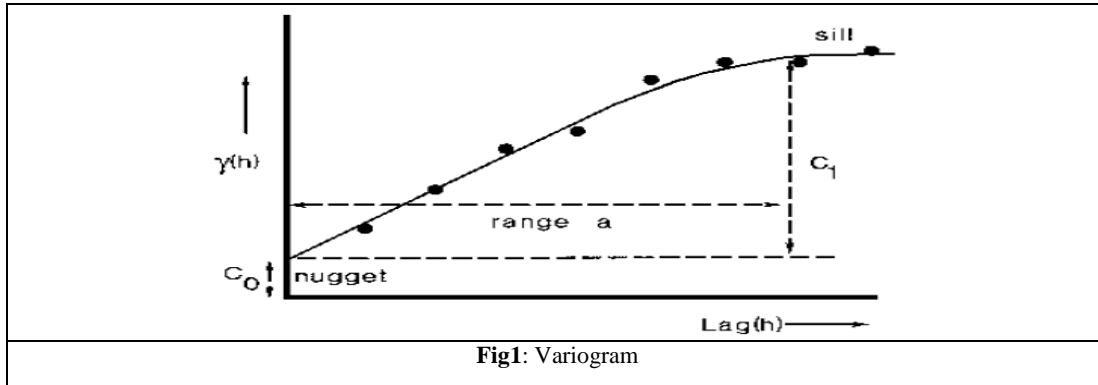
where  $Z(s)$  is the phenomena to be predicted at location  $s$  and  $\mu$  is a mean (structural function describing the structural component) which is known or unknown and it is either constant or varies with location (in data with trends), and these properties depends on the type of the used Kriging.  $\varepsilon(s)$  is the stochastic but spatially auto-correlated residuals from  $\mu(s)$  (i.e. the regionalized variable) and  $m$  is a constant accounts for un-biasness and can be defined as random noise having a normal distribution with a mean of 0 and a variance  $\sigma^2$ .

Based on the previous equation three main tasks are needed to perform geo-statistical modeling. The first task, is the determination of  $\mu(s)$  or trends in data followed by data de-trending through transformation techniques. This de-trending fulfills the assumption "Geo-statistics is a random process with dependence". Second task, is using the de-trended data twice: to uncover the dependency rules (estimate the spatial autocorrelation or get a function for the regionalized variable) and to make predictions using generalized linear regression techniques (kriging) of unknown values. The constant in this linear regression is a matrix calculated based on the spatial autocorrelation calculated for the sample points of known data. The spatial autocorrelation could be represented by the semi-variogram [10]

The semivariance can be estimated from the sample data using the formula:

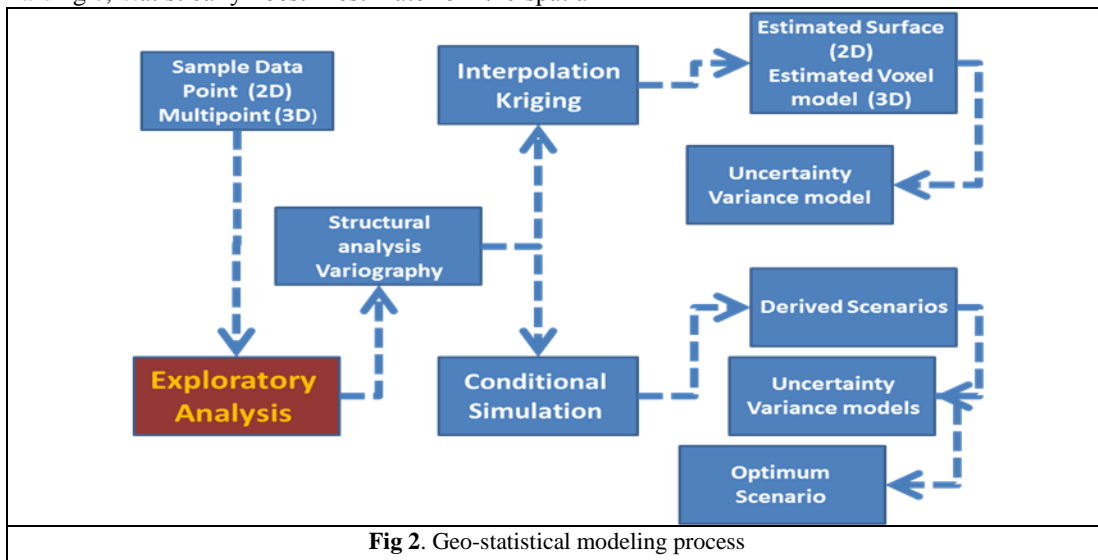
$$\hat{\gamma}(h) = \frac{1}{2n} \sum_{i=1}^n \{z(s) - z(s+h)\}^2 \quad [2].$$

where  $n$  is the number of pairs of sample points separated by distance  $h$ . The semivariance can be calculated for different values of  $h$  and plotted in what is referred to as an experimental variogram.



Variograms can be used in representing the regionalized variable in case they fulfill the concept of intrinsic stationarity which is the assumption that the variance of the difference is the same between any two points in the area of study that are at the same distance and direction apart no matter which two points are chosen. [6]. In other words, the semivariance depends only on the distance between the two points whatever their locations. In Geo-statistics there are two basic forms of prediction: estimation and simulation. In estimation, a single, statistically "best" estimate of the spatial

occurrence is produced. The estimation is based on both the sample data and on the variogram model. On the other hand, in simulation, many estimate scenarios (sometimes called "images") of the property distribution are produced, using the same model of spatial correlation as required for kriging. Differences between the alternative maps provide a measure of quantifying the uncertainty, an option not available with kriging estimation. [11] Fig 2. shows the geo-statistical modeling process for both estimation and simulation



### 3. How geo-statistics is applied in geotechnical engineering

#### Site Characterization.

The main challenge in site characterization is how to reveal the nature of soil profile just using a number of representative boreholes taking in consideration the uncertainties in soil formation and behavior. This needs a tool that can do two tasks. First, it needs to be able to estimate the whole soil profile based on the boreholes, Second, these tools needs to be able to evaluate the degree of confidence in the predicted soil profile. Geo-statistics can offer the two tasks through its two forms of prediction, estimation and

simulation. Standard deviation grids and cross validation at non-sampled location are considered indicators for the degree of confidence in the results and the capability of the model to estimate the values at these unsampled locations. Hence, it could be said that geo-statistics can optimize the site characterization process in three directions:

1. Application of the random process with dependence (that matches the heterogeneous nature of soil) to uncover the soil through the minimum number of boreholes.
2. Minimize the estimation errors that can result from the modeling process and the

type of the chosen model. This is by the application of the cross validation process

3. Direct the soil modeler to the locations that shows high standard deviation and need to carry out additional boreholes.

Moreover Geo-statistics can account for anomalies through a number of tools called data exploratory tools like Voronoi maps and QQ plots [6]. These anomalies determination tools can help in discovering data errors before the modeling process. While some times these anomalies could have significance specially when geo-statistics is used to determine the location of an Ore or a fault in rock mass.

The mostly applied geo-statistical interpolator in site characterization and soil mass profile estimation is indicator kriging. It is either used as an interpolator for estimation and mostly used in the simulation form through a technique called Sequential Indicator Simulation (SIS). It is usually used in the creation of the Voxel<sup>1</sup> 3D model of soil profile. [13]

(Hosseini, A.H etal 2006) [R15] has applied (SIS) in CPT data analysis to determine the site lithology. D. Folle etal 2006 [17] [18] applied geo-statistical techniques on more than 6000 boreholes archived in a GIS database to develop an integrated geological-geotechnical model of the subsoil of Rome.

#### **Other Geo-technical applications**

CHIN G. etal 2008 [19] and H.Kessler etal 2009 [20] have used geo-statistics in liquefaction analysis. M.B.Jaksa etal 1993 [R21], introduced Geo-statistics as a technique in the assessment of the range of correlation of the un-drained shear strength of clays.

It is worth mentioning that one of the main reasons that geo-statistics is not widely used in the field of geotechnical engineering is that there is not any geotechnical software package which is based on geo-statistics. Nearly all these packages are based on finite element analysis or empirical equations. Hence, it is aimed to show in this research a technique of implementing geo-statistics in geotechnical engineering through the integration between specialized geo-statistical software (Sgems) and a GIS software (ArcGIS)

#### **4. Sgems (Stanford Geo-statistical Modeling Software)**

The idea of Sgems started in 2001 when an Msc student "Nicolas Remy" introduced a C++ library with a title "GsTL:THE GEOSTATISTICAL

TEMPLATE LIBRARY IN C++" [1]. The main target of GsTL was to build a genuine programming library of geo-statistical tools and algorithms which is reusable in the production of new Geo-statistics products and could be imbedded in existing products. Remy was trying to overcome the limitations of GSLIB<sup>2</sup>, by being only a group of executable files without a mean of integrating these files with other programs. Hence, he produced the GsTL keeping in mind that the new library should be usable both for research developments and direct applications, should allow a fast reuse of existing code, should be easily extendable, should be understandable without too much computer science background. The GsTL library was developed using a technique of programming known as Generic programming that guarantees a degree of abstraction more effective than the well known Object oriented technique. In 2004, Remy applied his GsTL library in the development of a user friendly interface for Geostatistical modeling which called "Sgems". Sgems has two types of files based on the purpose of these files. First, it has the object files that are used for the definitions of the inputs and the outputs to the software. The Second type of files is the parameters files which are used to store the parameters needed for any of Sgems operations like variography, interpolation, simulation, data exploration, post processing, .....etc. He used two types of geometry objects which are the grids to represent the boundaries and the outputs of the geo-statistical interpolation process and the second type is the pointset which is used to store 3D points like boreholes. Whatever the object type, he used two file formats to store those objects which are the Ascii GSLIB format which is the most suitable for integration and another native binary format.

---

<sup>1</sup> A voxel : The 'voxel' term had originated from two words 'volume' and 'pixel'. A voxel is a volume element . Voxels are introduced to represent geo-bodies. The simplest representation of voxels has face that its shape is square. In the complex representation, face has a different size and shape [R14]

---

<sup>2</sup> GSLIB (Geo-statistical Software Library) is considered the first mature publicly available programming effort [R2] GSLIB, as its name suggests, is a collection of software, not a programming library: it provides a variety of computer executables which implement a broad family of algorithms, but it hardly provides a framework or tools for programming new software. [R1]

## 5. ArcGIS

The main target of using ArcGIS is to use its capabilities in Geospatial data handling considering, production, verification, manipulation, processing, presentation, spatial and attribute querying and reporting. Those capabilities are nearly missing in Sgems despite its high degree of maturity in Geo-statistical modeling. As a GIS product, ArcGIS can deal with both raster and vector data, so it can be used not only to represent the boreholes point set or interpolated grids as Sgems, but it can give a complete representation for the whole working site elements that could be needed for the modeling process. But still ArcGIS is missing the capability of dealing with real 3D Voxel data and this is one of the reasons for not using its Geo-statistical module in the modeling process. This deficit in ArcGIS was

## 6. Means of integration and applied technique

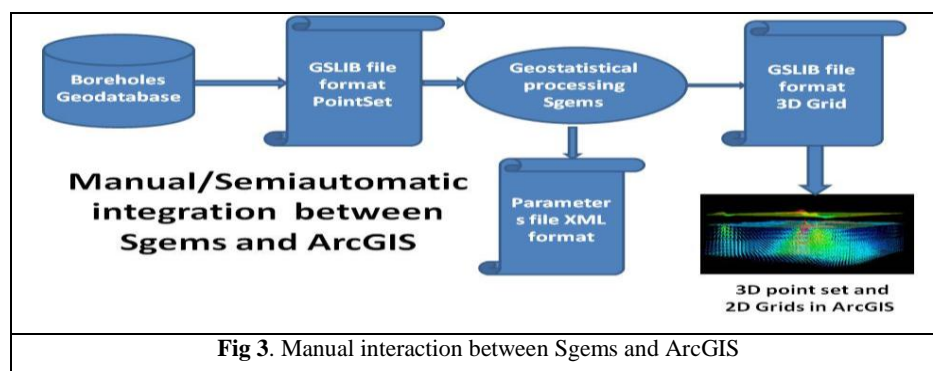
There are three means of integration between ArcGIS and Sgems for soil profile modeling:[ 14]

- Manual/Semi-automated integration depending on GSLIB Ascii files and Param XML Ascii files
- Call a program (like ArcGIS) from SGeMS through Python and C++ plug-in and use Sgems interface as the main working interface
- Call SGeMS externally (e.g. ArcGIS) and let Sgems work from behind as if it is a service not an application (Through Python scripts)

In the case study, the first integration technique was adopted. It depends on the nature of the GSLIB and Param files and the ability to create and edit them through delimited text editors like notepad or Excel. The process of integration started by the conversion of the source borehole data to Ascii file with the GSLIB format previously discussed. This conversion depends on the package incubating the source data which in the case of this study was ArcGIS software. This conversion could be done through some tools in ArcGIS in addition to delimited text editor like Excel or it could be done automatically through a program

covered in some ArcGIS third party applications (ArcGIS extensions) like Target of Geosoft Co., ArcHydro subsurface analyst of Aquaveo Co., EnterVol of Ctech and Rockworks of Rockware Co., but still some of these extensions do not support Geo-statistical algorithms like Rockworks or they are not supporting all the geo-statistical algorithms needed for different types of data like the case of EnterVol. For the main Geo-statistical extension of ArcGIS, it can deal only with 2D and not supporting the categorical data simulation algorithms which are the most suitable for boreholes modeling. Regarding automation, integration and processing, ArcGIS supports Python, .Net and any COM or .Net compliant programming. Hence, it can easily integrate with Sgems through Python.

with a programming language that depends on the package carrying the source data which in the case of ArcGIS Python or .Net or any other compliant language. After this conversion the GSLIB file became readable to the Sgems and it was able to read it as grid or pointset (which is the one used in the case of boreholes). The file was then used in interpolation process (Estimation or simulation). The parameters file of the interpolation process could be also produced using the text editor, but mainly it was generated through the Sgems interface due to the fact of needing manual intervention and evaluation in the determination of variograms parameters. The final output grid (as a result of interpolation and post processing) was saved as GSLIB file and was manipulated to be read by ArcGIS. This manipulation could be manually through ArcGIS tools and text editor or could be through a program. Finally, this manipulated file was presented as 3D points in ArcGIS. It is worth considering, that the process discussed in this section is not considering the ArcGIS third party programs mentioned above which enables ArcGIS to represent data in 3D voxel representation



### 7. Case study

Geo-statistical processing using the described technique of integration between ArcGIS and Sgems was applied on CPT data of Soil Improvement Works for Jaber Al Ahmed City Housing Project in Kuwait located approximately 25 km west of Kuwait's central area [16] to give a 3D model for the soil profile. The data (cone resistance & friction ratio graphs) of 30 CPT logs in section 2-4-3 of the project was converted to digital Excel table [R22] and Robertson chart [23] represented on ArcGIS

software [24] was used to classify the soil based on their cone resistance and friction ratio. The classified points were geographically represented using ArcGIS software. "SGems" [25] was applied on the geographically represented data to generate the 3D model of the soil profile using Sequential indicator simulation Geo-statistical algorithm. Finally, the 3D model was converted to a 3D point cloud to represent using the ArcScene interface of the ArcGIS software and to make spatial and attribute queries to deal with the generated model.

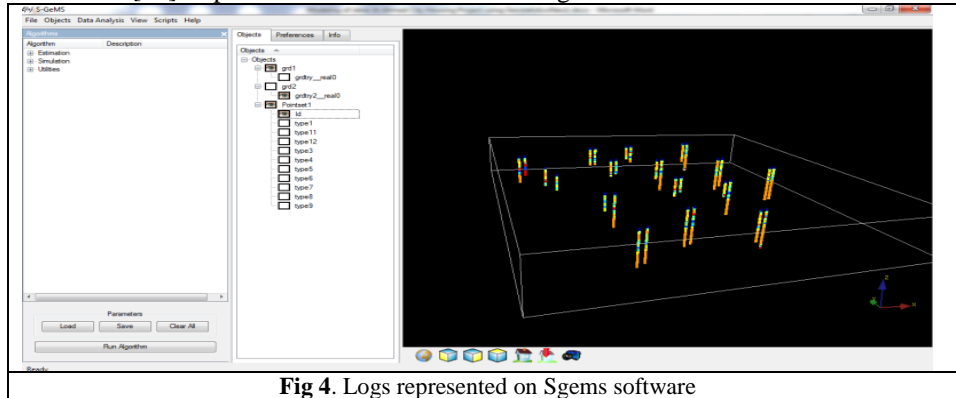


Fig 4. Logs represented on Sgems software

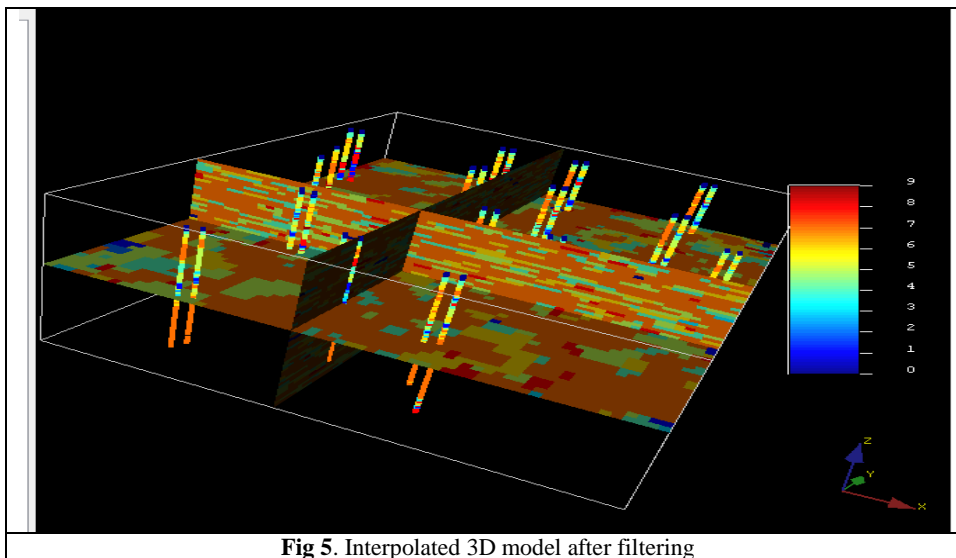


Fig 5. Interpolated 3D model after filtering

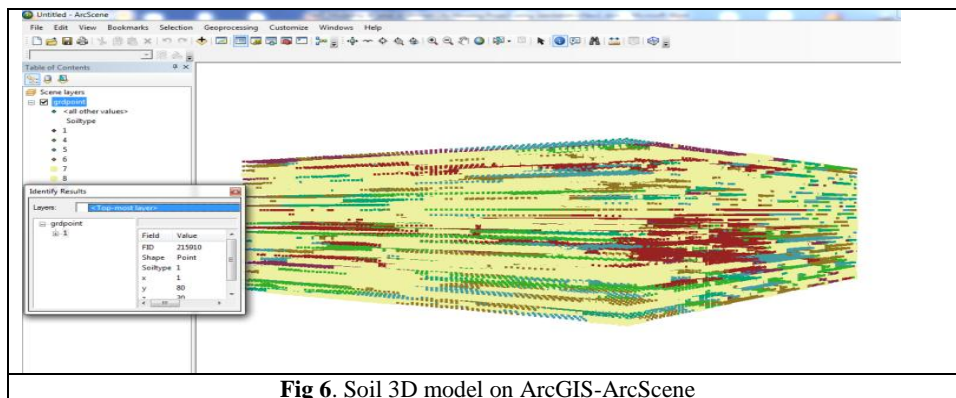


Fig 6. Soil 3D model on ArcGIS-ArcScene

## 8. Conclusion and recommendation

Geo-statistics is a powerful tool for geotechnical engineering, but it is still sidelined and needs more investigation from geotechnical engineers. The production of a specialized package for geotechnical engineering based on geo-statistics could be a solution to get benefit of these tools. Another approach could be through the integration between some known software for the geotechnical engineer and the specialized geo-statistical software. Geo-statistics as a technique is able to account for uncertainty in geotechnical modeling and it overcomes the assumptions of empirical equations used in traditional soil mechanics equations. Moreover, they are only limited to distance as a parameter for correlation which could minimize to some extent the problems that may arise when considering so many or inconvenient parameters like in the case of data mining. It is still needed to study more the new extensions of ArcGIS that could be used for 3D soil profile modeling using Geo-statistics. For future work, a fully automated process for soil profile modeling could be developed through the integration between Sgems and ArcGIS using Python or C++ or any COM compliant language

## References

- [1] Remy, N., 2001. GsTL: The Geostatistical Template Library in c++, MSc., Department of petroleum engineering of Stanford university.
- [2] Deutsch, C. and Journel, A. G., 1992. GSLIB: Geostatistical Software Library and User's Guide, Oxford University Press.
- [3] Das, B., 2006. Principles of Geotechnical Engineering., Thomson Learning.
- [4] De Weck, O. and Kim, Y., 2004. Finite Element Method, Engineering Design and Rapid Prototyping , Lectures, Massachusetts institute of technology
- [5] Giolas, A., 1994. A knowledge-based System For The Estimation of Geotechnical Properties., Phd., Durham University.
- [6] Johnston, K., Ver Hoef, J., Krivoruchko, K., and Lucas, N., 2003. ArcGIS® 9 Using ArcGIS® Geostatistical Analyst, ESRI.
- [7] Abdelfadil, M., 2009. Modeling Earth and Rockfill dams using GIS, MSc., AinShams university
- [8] Cressie, N., 1990. The Origins of Kriging, Mathematical Geology, v. 22, pp 239–252
- [9] Isaaks E. and Srivastava, R., 1989. An Introduction to Applied Geostatistics, Oxford University Press.
- [10] ESRI, 2013. <http://resources.arcgis.com/en/help/main/10.1>
- [11] Zhang, Y. 2011, Introduction to Geostatistics | Course Notes, Dept. of Geology & Geophysics, University of Wyoming
- [12] Hammah, R. and Curran J., 2003. Geostatistics in Geotechnical Engineering: Fad or Empowering, Article, RocNews Spring.
- [13] Pendrel, J., Leggett, M. and Mesdag, P., 2004. Geostatistical Simulation for Reservoir Characterization , CSGE National convention, Calgary, Canada.
- [14] Gupta R., 2010. New Features in SGeMS, Annual Meeting, Stanford Center for Reservoir Forecasting.
- [15] Hosseini, A., Biggar, K., Deutsch, C. and Mendoza, C., 2006. Geostatistical Analysis of CPT-UVIF Data for Development of a Site Conceptual Model, University of Alberta, Edmonton, Canada.
- [16] IncoLab 2011. Report about CPT results for Jaber Al-hamed City, Kuwait.
- [17] Folle, D., Raspa, G., Mancini, M., Moscatelli, M., Patera, A., Stigliano, F., Vallone R., Cavinato, G., Cavaretta, G., Milli, S., Garbin, F. and Storoni S., 2006. Geotechnical Modeling of The Subsoil of Rome (Italy) by Means of Multivariate Geostatistics , Int. Assoc. for Mathematical Geology XIth International Congress, Université de Liège, Belgium.
- [18] Folle, D., 2009. Application of Geotechnical Analysis on Urban Geologic - Geotechnical, Phd, Federal university of Do Roi Grande Do SUL
- [19] Chin, G., Tang, H., and Kan, S., 2008. Kriging Interpolation Method and 2.5D GIS Applied in the Probabilistic Estimation of Seismic Site Liquefaction, The 12th International Conference of International Association for Computer Methods and Advances in Geomechanics (IACMAG), Goa, India.
- [20] Kessler, H., Turner, A., Culshaw, M. and Royse, K., 2008. Unlocking the Potential of Digital 3D Geological Subsurface Models for Geotechnical Engineers, European conference of the International Association for Engineering geology, Madrid, Spain.
- [21] Jaksza, M., Kaggwa W. and Brooker, P., 1993. Geostatistical Modelling of the Spatial Variation of the Shear Strength of a Stiff, Overconsolidated Clay, Probabilistic Methods in Geotechnical Engineering, Li & Lo (eds), Balkema, Rotterdam, ISBN 90 5410 303 5.
- [22] Norman, G., 2013. <http://getdata-graph-digitizer.findmysoft.com/>
- [23] Robertson, P.K., 1990. Soil Classification Using the Cone Penetration Test, Canadian Geotechnical Journal, Edition 27(1), pp151-158.
- [24] Booth, B. and Mitchell, A., 2001. Getting Started with ArcGIS, ESRI press.
- [25] Remy, N., Boucher, A. and Wu J., 2011. Applied Geostatistics with SGeMS: A User's Guide, Amazon.